



RECEIVED

MAY 08 2002

TECH CENTER 1600/2900

SEQUENCE LISTING

<110> Max-Planck-Gesellschaft e.V.

A1 <120> Protein expression and structure solution using specific fusion vectors

<130> ST010209-EPA

<140> 01100762.2

<141> 2001-01-12

<160> 3

<170> PatentIn Ver. 2.1

<210> 1

<211> 765

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Partial myosin sequence of Dictyostelium; Component (1) of the recombinant protein M761-2R R238E

<400> 1

Met Asp Gly Thr Glu Asp Pro Ile His Asp Arg Thr Ser Asp Tyr His
1 5 10 15

Lys Tyr Leu Lys Val Lys Gln Gly Asp Ser Asp Leu Phe Lys Leu Thr
20 25 30

Val Ser Asp Lys Arg Tyr Ile Trp Tyr Asn Pro Asp Pro Lys Glu Arg
35 40 45

Asp Ser Tyr Glu Cys Gly Glu Ile Val Ser Glu Thr Ser Asp Ser Phe
50 55 60

Thr Phe Lys Thr Val Asp Gly Gln Asp Arg Gln Val Lys Lys Asp Asp
65 70 75 80

Ala Asn Gln Arg Asn Pro Ile Lys Phe Asp Gly Val Glu Asp Met Ser
85 90 95

Glu Leu Ser Tyr Leu Asn Glu Pro Ala Val Phe His Asn Leu Arg Val
100 105 110

Part of
paper No. 3

RECEIVED

APR 04 2002

TECH CENTER 1600/2900

RECEIVED

APR 04 2002

TECH CENTER 1600/2900

Arg Tyr Asn Gln Asp Leu Ile Tyr Thr Tyr Ser Gly Leu Phe Leu Val
115 120 125

Ala Val Asn Pro Phe Lys Arg Ile Pro Ile Tyr Thr Gln Glu Met Val
130 135 140

Asp Ile Phe Lys Gly Arg Arg Arg Asn Glu Val Ala Pro His Ile Phe
145 150 155 160

(W)
Ala Ile Ser Asp Val Ala Tyr Arg Ser Met Leu Asp Asp Arg Gln Asn
165 170 175

Gln Ser Leu Leu Ile Thr Gly Glu Ser Gly Ala Gly Lys Thr Glu Asn
180 185 190

Thr Lys Lys Val Ile Gln Tyr Leu Ala Ser Val Ala Gly Arg Asn Gln
195 200 205

Ala Asn Gly Ser Gly Val Leu Glu Gln Gln Ile Leu Gln Ala Asn Pro
210 215 220

Ile Leu Glu Ala Phe Gly Asn Ala Lys Thr Thr Arg Asn Asn Asn Ser
225 230 235 240

Ser Arg Phe Gly Lys Phe Ile Glu Ile Gln Phe Asn Ser Ala Gly Phe
245 250 255

Ile Ser Gly Ala Ser Ile Gln Ser Tyr Leu Leu Glu Lys Ser Arg Val
260 265 270

Val Phe Gln Ser Glu Thr Glu Arg Asn Tyr His Ile Phe Tyr Gln Leu
275 280 285

Leu Ala Gly Ala Thr Ala Glu Glu Lys Lys Ala Leu His Leu Ala Gly
290 295 300

Pro Glu Ser Phe Asn Tyr Leu Asn Gln Ser Gly Cys Val Asp Ile Lys
305 310 315 320

Gly Val Ser Asp Ser Glu Glu Phe Lys Ile Thr Arg Gln Ala Met Asp
325 330 335

Ile Val Gly Phe Ser Gln Glu Glu Gln Met Ser Ile Phe Lys Ile Ile
340 345 350

Ala Gly Ile Leu His Leu Gly Asn Ile Lys Phe Glu Lys Gly Ala Gly
355 360 365

Glu Gly Ala Val Leu Lys Asp Lys Thr Ala Leu Asn Ala Ala Ser Thr
370 375 380

Val Phe Gly Val Asn Pro Ser Val Leu Glu Lys Ala Leu Met Glu Pro
385 390 395 400

Arg Ile Leu Ala Gly Arg Asp Leu Val Ala Gln His Leu Asn Val Glu
405 410 415

Lys Ser Ser Ser Ser Arg Asp Ala Leu Val Lys Ala Leu Tyr Gly Arg
420 425 430

Leu Phe Leu Trp Leu Val Lys Lys Ile Asn Asn Val Leu Cys Gln Glu
435 440 445

Arg Lys Ala Tyr Phe Ile Gly Val Leu Asp Ile Ser Gly Phe Glu Ile
450 455 460

Phe Lys Val Asn Ser Phe Glu Gln Leu Cys Ile Asn Tyr Thr Asn Glu
465 470 475 480

Lys Leu Gln Gln Phe Phe Asn His His Met Phe Lys Leu Glu Gln Glu
485 490 495

Glu Tyr Leu Lys Glu Lys Ile Asn Trp Thr Phe Ile Asp Phe Gly Leu
500 505 510

Asp Ser Gln Ala Thr Ile Asp Leu Ile Asp Gly Arg Gln Pro Pro Gly
515 520 525

Ile Leu Ala Leu Leu Asp Glu Gln Ser Val Phe Pro Asn Ala Thr Asp
530 535 540

Asn Thr Leu Ile Thr Lys Leu His Ser His Phe Ser Lys Lys Asn Ala
545 550 555 560

Lys Tyr Glu Glu Pro Arg Phe Ser Lys Thr Glu Phe Gly Val Thr His
565 570 575

Tyr Ala Gly Gln Val Met Tyr Glu Ile Gln Asp Trp Leu Glu Lys Asn
580 585 590

Lys Asp Pro Leu Gln Gln Asp Leu Glu Leu Cys Phe Lys Asp Ser Ser
595 600 605

Asp Asn Val Val Thr Lys Leu Phe Asn Asp Pro Asn Ile Ala Ser Arg
610 615 620

Ala Lys Lys Gly Ala Asn Phe Ile Thr Val Ala Ala Gln Tyr Lys Glu
625 630 635 640

Gln Leu Ala Ser Leu Met Ala Thr Leu Glu Thr Thr Asn Pro His Phe
645 650 655

Val Arg Cys Ile Ile Pro Asn Asn Lys Gln Leu Pro Ala Lys Leu Glu
660 665 670

A1 Asp Lys Val Val Leu Asp Gln Leu Arg Cys Asn Gly Val Leu Glu Gly
675 680 685

Ile Arg Ile Thr Arg Lys Gly Phe Pro Asn Arg Ile Ile Tyr Ala Asp
690 695 700

Phe Val Lys Arg Tyr Tyr Leu Leu Ala Pro Asn Val Pro Arg Asp Ala
705 710 715 720

Glu Asp Ser Gln Lys Ala Thr Asp Ala Val Leu Lys His Leu Asn Ile
725 730 735

Asp Pro Glu Gln Tyr Arg Phe Gly Ile Thr Lys Ile Phe Phe Arg Ala
740 745 750

Gly Gln Leu Ala Arg Ile Glu Glu Ala Arg Glu Gln Arg
755 760 765

<210> 2

<211> 1016

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Whole sequence
of recombinant protein M761-2R R238 E

<220>

<223> The protein comprises as component (1) the aa
sequence of myosin II motor domain of
Dictyostelium, a three aa linker region and the
a-actinin aa sequence

<400> 2

Met Asp Gly Thr Glu Asp Pro Ile His Asp Arg Thr Ser Asp Tyr His
1 5 10 15

Lys Tyr Leu Lys Val Lys Gln Gly Asp Ser Asp Leu Phe Lys Leu Thr

20

25

30

Val Ser Asp Lys Arg Tyr Ile Trp Tyr Asn Pro Asp Pro Lys Glu Arg
35 40 45

Asp Ser Tyr Glu Cys Gly Glu Ile Val Ser Glu Thr Ser Asp Ser Phe
50 55 60

Thr Phe Lys Thr Val Asp Gly Gln Asp Arg Gln Val Lys Lys Asp Asp
65 70 75 80

Ala Asn Gln Arg Asn Pro Ile Lys Phe Asp Gly Val Glu Asp Met Ser
85 90 95

Glu Leu Ser Tyr Leu Asn Glu Pro Ala Val Phe His Asn Leu Arg Val
 100 105 110

Arg Tyr Asn Gln Asp Leu Ile Tyr Thr Tyr Ser Gly Leu Phe Leu Val
 115 120 125

Ala Val Asn Pro Phe Lys Arg Ile Pro Ile Tyr Thr Gln Glu Met Val
130 135 140

Asp Ile Phe Lys Gly Arg Arg Arg Asn Glu Val Ala Pro His Ile Phe
145 150 155 160

Ala Ile Ser Asp Val Ala Tyr Arg Ser Met Leu Asp Asp Arg Gln Asn
165 170 175

Gln Ser Leu Leu Ile Thr Gly Glu Ser Gly Ala Gly Lys Thr Glu Asn
180 185 190

Thr Lys Lys Val Ile Gln Tyr Leu Ala Ser Val Ala Gly Arg Asn Gln
165

Ala Asn Gly Ser Gly Val Leu Glu Gln Gln Ile Leu Gln Ala Asn Pro
G16

Ile Leu Glu Ala Phe Gly Asn Ala Lys Thr Thr Arg Asn Asn Asn Ser
225

Ser Arg Phe Gly Lys Phe Ile Glu Ile Gln Phe Asn Ser Ala Gly Phe

Ile Ser Gly Ala Ser Ile Gln Ser Tyr Leu Leu Glu Lys Ser Arg Val

Val Phe Glu Ser Glu Thr Gly Pro Ile Asp

275

280

285

Leu Ala Gly Ala Thr Ala Glu Glu Lys Lys Ala Leu His Leu Ala Gly
290 295 300

Pro Glu Ser Phe Asn Tyr Leu Asn Gln Ser Gly Cys Val Asp Ile Lys
305 310 315 320

Gly Val Ser Asp Ser Glu Glu Phe Lys Ile Thr Arg Gln Ala Met Asp
325 330 335

R1 Ile Val Gly Phe Ser Gln Glu Glu Gln Met Ser Ile Phe Lys Ile Ile
340 345 350

Ala Gly Ile Leu His Leu Gly Asn Ile Lys Phe Glu Lys Gly Ala Gly
355 360 365

Glu Gly Ala Val Leu Lys Asp Lys Thr Ala Leu Asn Ala Ala Ser Thr
370 375 380

Val Phe Gly Val Asn Pro Ser Val Leu Glu Lys Ala Leu Met Glu Pro
385 390 395 400

Arg Ile Leu Ala Gly Arg Asp Leu Val Ala Gln His Leu Asn Val Glu
405 410 415

Lys Ser Ser Ser Arg Asp Ala Leu Val Lys Ala Leu Tyr Gly Arg
420 425 430

Leu Phe Leu Trp Leu Val Lys Lys Ile Asn Asn Val Leu Cys Gln Glu
435 440 445

Arg Lys Ala Tyr Phe Ile Gly Val Leu Asp Ile Ser Gly Phe Glu Ile
450 455 460

Phe Lys Val Asn Ser Phe Glu Gln Leu Cys Ile Asn Tyr Thr Asn Glu
465 470 475 480

Lys Leu Gln Gln Phe Phe Asn His His Met Phe Lys Leu Glu Gln Glu
485 490 495

Glu Tyr Leu Lys Glu Lys Ile Asn Trp Thr Phe Ile Asp Phe Gly Leu
500 505 510

Asp Ser Gln Ala Thr Ile Asp Leu Ile Asp Gly Arg Gln Pro Pro Gly
515 520 525

Ile Leu Ala Leu Leu Asp Glu Gln Ser Val Phe Pro Asn Ala Thr Asp

530

535

540

Asn Thr Leu Ile Thr Lys Leu His Ser His Phe Ser Lys Lys Asn Ala
545 550 555 560

Lys Tyr Glu Glu Pro Arg Phe Ser Lys Thr Glu Phe Gly Val Thr His
565 570 575

Tyr Ala Gly Gln Val Met Tyr Glu Ile Gln Asp Trp Leu Glu Lys Asn
580 585 590

Lys Asp Pro Leu Gln Gln Asp Leu Glu Leu Cys Phe Lys Asp Ser Ser
595 600 605

Asp Asn Val Val Thr Lys Leu Phe Asn Asp Pro Asn Ile Ala Ser Arg
610 615 620

Ala Lys Lys Gly Ala Asn Phe Ile Thr Val Ala Ala Gln Tyr Lys Glu
625 630 635 640

Gln Leu Ala Ser Leu Met Ala Thr Leu Glu Thr Thr Asn Pro His Phe
645 650 655

Val Arg Cys Ile Ile Pro Asn Asn Lys Gln Leu Pro Ala Lys Leu Glu
660 665 670

Asp Lys Val Val Leu Asp Gln Leu Arg Cys Asn Gly Val Leu Glu Gly
675 680 685

Ile Arg Ile Thr Arg Lys Gly Phe Pro Asn Arg Ile Ile Tyr Ala Asp
690 695 700

Phe Val Lys Arg Tyr Tyr Leu Leu Ala Pro Asn Val Pro Arg Asp Ala
705 710 715 720

Glu Asp Ser Gln Lys Ala Thr Asp Ala Val Leu Lys His Leu Asn Ile
725 730 735

Asp Pro Glu Gln Tyr Arg Phe Gly Ile Thr Lys Ile Phe Phe Arg Ala
740 745 750

Gly Gln Leu Ala Arg Ile Glu Glu Ala Arg Glu Gln Arg Leu Gly Ser
755 760 765

Glu Gln Thr Lys Ser Asp Tyr Leu Lys Arg Ala Asn Glu Leu Val Gln
770 775 780

Trp Ile Asn Asp Lys Gln Ala Ser Leu Glu Ser Arg Asp Phe Gly Asp

785 790 795 800
Ser Ile Glu Ser Val Gln Ser Phe Met Asn Ala His Lys Glu Tyr Lys
805 810 815

Lys Thr Glu Lys Pro Pro Lys Gly Gln Glu Val Ser Glu Leu Glu Ala
820 825 830

Ile Tyr Asn Ser Leu Gln Thr Lys Leu Arg Leu Ile Lys Arg Glu Pro
835 840 845

Phe Val Ala Pro Ala Gly Leu Thr Pro Asn Glu Ile Asp Ser Thr Trp
850 855 860

Ser Ala Leu Glu Lys Ala Glu Gln Glu His Ala Glu Ala Leu Arg Ile
865 870 875 880

Glu Leu Lys Arg Gln Lys Lys Ile Ala Val Leu Leu Gln Lys Tyr Asn
885 890 895

Arg Ile Leu Lys Lys Leu Glu Asn Trp Ala Thr Thr Lys Ser Val Tyr
900 905 910

Leu Gly Ser Asn Glu Thr Gly Asp Ser Ile Thr Ala Val Gln Ala Lys
915 920 925

Leu Lys Asn Leu Glu Ala Phe Asp Gly Glu Cys Gln Ser Leu Glu Gly
930 935 940

Gln Ser Asn Ser Asp Leu Leu Ser Ile Leu Ala Gln Leu Thr Glu Leu
945 950 955 960

Asn Tyr Asn Gly Val Pro Glu Leu Thr Glu Arg Lys Asp Thr Phe Phe
965 970 975

Ala Gln Gln Trp Thr Gly Val Lys Ser Ser Ala Glu Thr Tyr Lys Asn
980 985 990

Thr Leu Leu Ala Glu Leu Glu Arg Leu Gln Lys Ile Glu Asp Ala Leu
995 1000 1005

His His His His His His His
1010 1015

&
<210> 3
<211> 3048
<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: DNA sequence
coding for recombinant protein M761-2R R238E

<400> 3

atggatgta ccgaggatcc aattcatgat agaacttcag attatcacaa atactaaaaa 60
gtttaaacaag gtgattctga ttttattaaa cttactgtt cagataagag atacatttg 120
tataatccag atccaaaaga aagagattca tatgaatgtg gtgaaattgt ttcagaaacc 180
tctgattctt tcacattcaa aaccgttgc ggtcaagaca gacaagtcaa aaaggatgtat 240
gccaatcaac gtaatccaat caaattcgat ggtgtcgaag atatgtctga attatcatac 300
ctcaatgaac cagcagttt ccacaatctc cgtgttcgtt acaatcaaga ttttaatttac 360
acctattcag gtctctttt gggtgccgtc aatccattca agagaattcc aatctacact 420
caagagatgg ttgatatctt caaaggcgtt agaagaaatg aagttgcccc acatatttc 480
gccatttctg atgtgccta tcgttcaatg tttagatgatc gtcaaaatca atcactctta 540
atcaactggtg aatctgggtc tggtaagact gaaaacacca aaaaggtcat tcaatatctt 600
gcatctgtcg ctggcgttaa tcaagccaat ggttaggtg tattggaca acaaattctc 660
caagccaatc caatccttga agcttttgtt aatgccaaaa ccacccgtaa caacaatca 720
tctcgtttcg gtaaattcat tgaaattcaa ttcaacagtg ctggttcat tagtgggtct 780
tcaattcaat cctacctttt agagaaatca cgtgtcgtt tccaatctga aaccgaacgt 840
aattatcaca ttttctatca actcttagct ggtgccaccc ccgaagaaaa gaaagctctt 900
caactggctg gtcagaatc attcaactac ttaaatcaaa gtgggtgtgt tgatataaaa 960
ggtgtctctg atagtgaaga attcaaaatc actcgtcaag ctatggacat tgggttttc 1020
tcacaagaag aacaatgtc aatctttaag atcattgctg gtatcttaca ttttaggttaac 1080
atcaaattcg aaaaagggtgc tggtaaggt gctgtcctca aagacaaaaac cgccctcaac 1140
gctgcttcaa ccgtcttgg tgtcaatcca tcagtccttgg aaaaggctct catggaaacca 1200
cgtatTTtag ccggcgtgta tttagttgtt caacatctca acgttggaaaa atcctcatca 1260
tcaagagacg ctcttgcata agctctctat ggtcgtctt tcctctgggtt ggtcaaaaaag 1320
atcaacaatg tcctctgtca agagagaaaa gcttacttta ttgggttttt ggatattca 1380
ggttttgaaa ttttcaaagt caattcattc gaacaattat gtatcaatta taccaatgaa 1440
aaactccaaac aattcttcaa tcaccatatg ttcaaattgg aacaagaaga atatctaaaa 1500
gagaaaaatca attggacttt catcgatTTT ggtcttgatt cacaagccac tatcgattta 1560
attgatggtc gtcaaccacc aggtatTTT gctctttgg atgaacaatc tgggttccca 1620
aatgccaccg ataatacttt aatcacaaaa ctccacagtc acttagcaa gaagaacgcc 1680
aaatacgaag aaccacgttt ctccaaaaacc gaattttgggtt ttaccattta tgctggtcaa 1740
gtcatgtatg agattcaaga ttggtagaa aagaacaaag atccattaca acaagatctc 1800
gaactttgtt tcaaagattc atcagacaac gttgtcacca aactttcaa tgatccaaac 1860
attgccagtc gtgcaaagaa aggtgcaaacc tttatcactg tcgcccgtca atacaaggaa 1920
caattagcct cactcatggc nacccttcaa accaccaacc cacatttcgt tcgttgtatc 1980
attccaaaca acaaacaatt accagccaaa ctcgaagata aagttgtcct cgaccaatta 2040
cggtgcaatg gtgtcctcgaa aggtattcgtt attactcgta aaggtttccc aaatcgatt 2100
atctatgccc atttcgtcaa acgttactat ttattagctc caaacgttcc aagagacgct 2160
gaagactcac aaaaagccac cgatgctgtt ctcaaacatc ttaacattga tccagaacaa 2220
tatcgtttcg gtatcaccaa gattttcttc cgtgccggtc aattagctcg tattgaagaa 2280
gctcgtgaac aacgtctagg atccgaacaa accaaatctg attatcttaa aagagccaaat 2340
gaactcggttc aatggattaa cgataaacaa gcatcacttg aatcacgtga ttttgggtat 2400
tccatcgaaat ctgttcaaag tttcatgaac gctcataaaag aatataaaaaa aaccgaaaaa 2460

A1

ccaccaaagg gtcaagaagt ctctgaattt gaagcttatct acaattcatt acaaactaaa 2520
ttacgtttaa ttaaacgtga accattgtt gcaccagctg gtctcaactcc aaatgaaatc 2580
gattccactt ggtccgcttt agagaaagct gaacaagaac atgctgaagc cctccgtatt 2640
gaactcaaac gtcaaaaagaa aattgcagtt ctcttacaaa aatacaatcg tattctcaag 2700
aaactcgaaa actgggccac caccaaatct gtctacctcg gttccaatga aaccggtgac 2760
agtatcaactg ctgttcaagc taaattaaag aatttagaag ctttgatgg tgaatgtcaa 2820
tcatttggaaag gtcaatcaaactctc ctctgatctc ctcagcatc ttgctcaatt aactgaactc 2880
aactacaatg gtgtaccaga actcaactgaa cgtaaagata cattcttgc tcaacaatgg 2940
actggtgtta aatcatctgc tgaaacctac aaaaacactc ttttagctga acttgaaaga 3000
ctccaaaaga ttgaagatgc attacatcat catcatcatc atcaccac 3048